MASKING EFFECTS OF PERCEPTUAL CUES ON HEARING-IMPAIRED EARS FOR SPEECH PERCEPTION

BY

SIYAN GUO

THESIS

Submitted in partial fulfillment of the requirements
for the degree of Master of Science in Electrical and Computer Engineering
in the Graduate College of the
University of Illinois at Urbana-Champaign, 2019

Urbana, Illinois

Adviser:

Professor Jont Allen

ABSTRACT

The thesis discusses a strategy for evaluating how the hearing-impaired (HI) ear decodes speech sounds by analyzing the impact of manipulating the perceptual speech cues (Li and Allen, 2011), followed by a confusion matrix (CM) error pattern analysis. This study analyzes subjects' responses based on consonant errors. Following this analysis, key insights are provided into the HI listening strategy.

Two experiments on consonant-vowel identification were conducted on five normal-hearing (NH) and ten HI subjects. Subjects' performance includes different primary cue intensities (i.e., signal-to-noise ratio) with and without conflicting cues (Kapoor and Allen, 2012). A classification scheme is then introduced to quantify HI ears based on both their error patterns and hearing loss profiles. The idiosyncratic nature of the HI ears is then studied and further quantified using an entropy metric.

This analysis demonstrates that: (1) HI ears are using the same primary cue as NH ears, and (2) some HI ears also use conflicting cues for consonant identification. These conclusions are unique to this study (and the unpublished PhD thesis of Cole (2017)).

This classification method should be clinically useful in predicting success with the HI aided condition, as well as in improving the insertion gain fitting procedure.

To Father and Mother, for their love and support.

ACKNOWLEDGMENTS

This work would not have been possible without the support of many people.

I am thankful for my adviser, Prof. Jont B. Allen, for his guidance on this project, and former PhD student Cliston L. Cole for performing the experiments and collecting the data.

I am grateful for my parents, who love me and encourage me, and who endured this long journey with me.

I am thankful for my friends, who have brought me much joy and company when I felt lonely.

TABLE OF CONTENTS

CHAPTER 1 INTRODUCTION 1

CHAPTER 2 METHODS 6

CHAPTER 3 EXPERIMENT I RESULTS 12

CHAPTER 4 EXPERIMENT II RESULTS 20

CHAPTER 5 ENTROPY ANALYSIS 25

CHAPTER 6 CONCLUSION 30

REFERENCES 32

CHAPTER 1

INTRODUCTION

The primary purpose of hearing aids is to improve speech perception for hearing-impaired (HI) ears. The HI often complain about the performance of hearing aids when the amplification does not help them understand speech in noise. Of course background noise can affect everyone, but people with moderate hearing loss are much more sensitive to noise. It is likely that this sensitivity results from selectively damaged cochlear critical bands, which have partially lost their ability to filter out noise. The loss of filtering ability increases the masking effect of noise on speech cues.

The characteristic of this masking on speech cues has not been demonstrated. This research shall show that the HI ear cannot understand speech in noise because it cannot decode critical speech cues, due to an interaction between its hearing loss and the masking effects of the noise on the cues. The key feature of this loss is its idiosyncratic distribution.

Numerous types of analog and digital hearing aids use a variety of signal processing algorithms to improve speech perception. Most if not all commercial devices use multi-band compression [1,2], and many use noise and feedback reduction techniques, adaptive directional signal processing, and environment classification [3]. However, in the personal experience of Cole, who collected the data, these signal processing algorithms do not significantly improve speech perception [4]. It is noted, however, that the success of these methods depends on the degree of loss (i.e., is highly idiosyncratic).

One of the most common problems of current hearing aids is their amplification strategy. For example, some hearing aids' amplification strategies amplify all high, mid, and low frequencies [e.g. National Acoustic Laboratories - Revised (NALR) [3–5]]. To make hearing aids that better enhance speech perception, it is important that designers of the signal processing algorithms precisely understand the necessary perceptual cues required for correct recognition in HI ears. This research shall show that the failure of past strategies is due to an insufficient understanding of critical speech cues and their sensitivity to masking noise.

Some studies have concluded that reduced audibility of speech cues, due to hearing loss and external noise, is sufficient to characterize the difficulties HI ears experience in speech perception [6]. Other studies have suggested that HI ears have an additional deficit, beyond audibility, defined as "SNR loss." A flaw in [6] is that they used an average signal-to-noise ratio (SNR) by averaging across all the sounds. All ears were treated the same, given the same loss. They did not consider the differences given a fixed hearing loss. Namely since they did not take into account the idiosyncratic listening behavior across different HI ears, they came to an incorrect conclusion. Furthermore, the errors depend strongly on the utterance (i.e., token) [7, 8]. For example the male token of /ba/ might have 100% error and the female token zero error at the same SNR. The error is highly dependent on the production of the token, with most tokens (80%) having zero error at -2 dB SNR [9–12]. Most of the errors in speech perception for normal hearing (NH) ears are due to production rather than perception.

The main research question being asked here is: How can we determine the strategy used by each HI ear when decoding consonants? Based on previous results, where the speech cues were masked in NH ears [8, 13, 14], we hypothesize that HI ears will show differences not predicted by their hearing loss profile.

HI ear confusions shall be explained based on these earlier primary cue masking results. Two studies on NH ears [13, 15] showed that the masking of primary cue, and/or removing the conflicting cues, can change and even improve speech perception (in NH ears). Given these new results in NH ears, we wish to explore the same effects of primary and conflicting cues in HI ears.

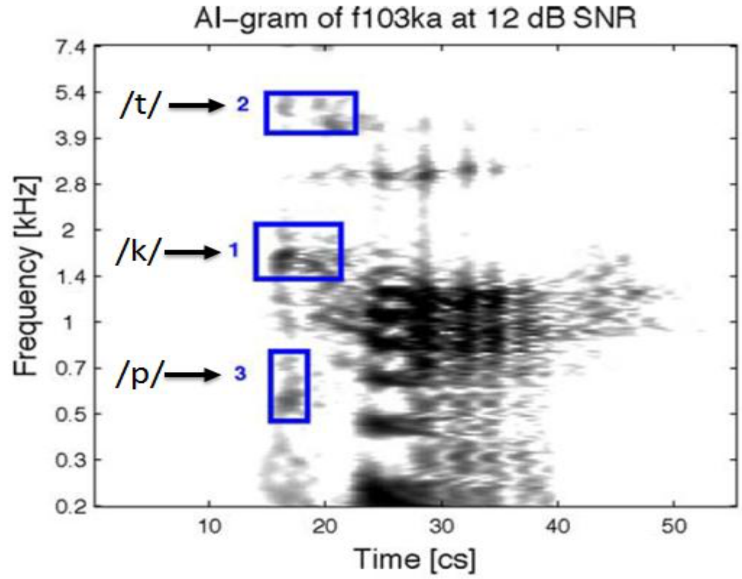


Figure 1.1: *AI gram of spoken utterance /ka/ by female talker 103 at 12 [dB] SNR. Three regions are labeled 1, 2, 3, where 1 is the primary cue and 2 and 3 are conflicting and possibly secondary cues, produced naturally along with 1. As best we know, only 2 contributes to the identification of /ka/ [16]. Figure adapted from Li and Allen [15].*

Namely the HI error patterns shall be generalized by examining errors due to the masking of the primary and conflicting cues. We hypothesize that this will lead to insights into the HI ear's phone decoding strategy. If patterns in the idiosyncratic HI phone decoding strategies can be determined, hearing aid designers will be able to develop better amplification strategies, and take advantage of idiosyncratic listening, by selectively reducing these errors.

It is essential to understand the basic concepts of perceptual cues used for speech perception. It would be of both theoretical and practical value to show that the HI ears use the same cues as NH ears. It seems unlikely that a person would switch cues as they began to lose their hearing. Little is known about the speech cues used by HI ears. This is especially true for cochlear implant subjects.

In the NH ear, primary perceptual cues are time-frequency energy patterns of spoken utterances [15–17]. Perceptual cues can be visualized with analytical tools such as the AIgram (a graphical version of the Articulation Index) [17]. An AIgram example for utterance /ka/ spoken by talker f103 is shown in Fig. 1.1. The patterns within the outline labeled 1 form the primary cue for /ka/. Any acoustic features outside the outlined primary cue are secondary cues. Labels 2 and 3, /ta/ and /pa/ acoustic cues, respectively, are a special type of secondary cue for /ka/, called conflicting cues, which are contrasting acoustic features to the primary cue. When a talker speaks a /ka/, the listener can be confused by the conflicting cues /ta/ or /pa/, especially in the presence of noise [17]. This effect was first explored by [13].

The primary cue is the critical information that the NH listener uses to correctly identify an utterance. This research shall show that HI ears use the same primary cue as the NH ears to make correct choices. The reasons the HI cannot understand speech in noise as well as NH ears are that (1) they cannot hear the subset of critical speech cues, due to both hearing loss and masking effects of the noise and (2) the conflicting cues act as the dominant source of confusion when the hearing loss renders the primary cue less intelligible.

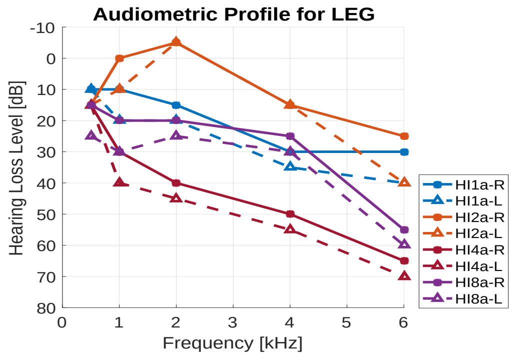
CHAPTER 2

METHODS

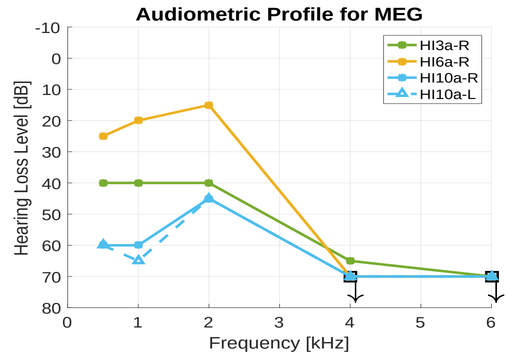
Two perceptual modification experiments were conducted in this study, with the goal of understanding the role of plosive cues in HI ears. First the degree of masking on the primary cue is studied. Second the conflicting cues are removed. An understanding of the role of these cues should help to explain the idiosyncratic error patterns made by each HI ear, as previously determined for NH ears [13].

A total of 5 NH and 10 HI subjects were recruited. The 5 NH subjects were averaged for analysis to form the ANH group. The 10 HI subjects expanded into $N = 17$ HI ears. All subjects ranged in age from 26 to 65 years old. A pure-tone audiometry (PTA) test was administered to each HI subject prior to speech testing. The PTA results are displayed in Fig. 2.1. As shown, the HI subjects separated into three well defined groups, as described in Chapter 3. The experiment was performed under University of Illinois IRB protocol [4].

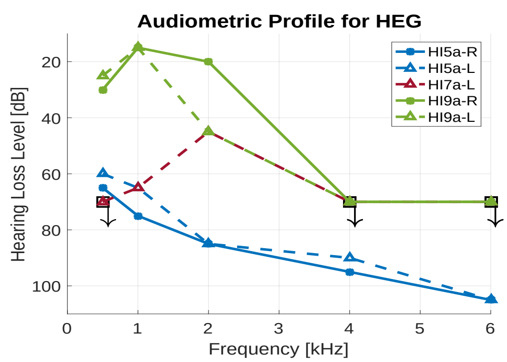
Based on the similarity of the five NH subjects, the responses were averaged together to increase the power in the NH results. This fictitious ANH subject responded



(a) *Audiogram for LEG*



(b) *Audiogram for MEG*



(c) *Audiogram for HEG*

Figure 2.1: *Hearing impaired audiometric profiles for the subjects in this study. Based on the results of the study the HI subjects have been divided into three severity groups: low, medium and high error groups, denoted LEG, MEG and HEG. Hearing loss levels for both ears for HI subjects are provided, but are not available for HI3a, HI6a, and HI7a. Subscript R/L indicates the right or left ear. The black downward pointing arrows indicate ears that have hearing loss greater than 70 [dB]. Hearing loss greater than 70 [dB] is considered severe. HI5a, HI6a, HI7a, HI9a, and HI10a have at least one ear with a severe hearing loss at 4 and 6 [kHz]. HI5a has profound hearing loss (i.e. $\geq$ 90 [dB]) for both ears at 4 and 6 [kHz].*

to pairs of four different consonants, spoken by two different talkers, resulting in 8 tokens. Each token was represented at three SNR levels, and each subject listened to each sound once ($N = 15$). With further averaging across three conditions (unmodified, +6 and +12), the total number of ANH trials per consonant increased to $N = 45$. There was exactly one error (token f103ka) in the unmodified condition, giving a final /ka/ error of $1/45 = 2.22$[%].

In 2007 an identical experiment (UIUCs04, aka MN64) was performed, except 10 different NH subjects were used [18]. In the 2007 experiment there was a ≈3[%] error at -10 dB SNR, and zero error at -2 [dB] and quiet (no added noise). By combining the two experiments, which only differed in the subjects, only the present experiment resulted in a single error above -2 [dB] SNR.

Given only 1 error for the four consonants for the three conditions and 15 NH subjects, the consonant error is statistically zero ($P_e \approx 0$). Thus it is reasonable to average the subject-conditions together. This justifies defining the fictitious "ANH subject" having zero error at and above 0 [dB] SNR.

Procedures: Each subject was seated in the sound booth, and began by listening to trials to adjust the speech to their most comfortable level (MCL). Prior to the experiment session, a short practice session was provided to allow the subject to become familiar with the displayed 20 options and the GUI interface, comprised of the 18 CV syllables, "Only Noise" and "Other" options. The subject had the option to replay the stimulus two additional times following each trial. All stimuli in the practice session were presented at 18 [dB] SNR. Following the practice session, the stimuli were presented at SNR levels of 0, 9, 18 [dB] [4, 13].

Four plosive consonants /t,k,d,g/ paired with vowel /a/ in CV context were used as target speech stimuli. Each CV was presented using two different talkers. Each talker and CV combination is denoted as a "token." There are a total of eight tokens of mixed gender in this experiment, with the exception of the two female /da/ consonants. An additional six non-plosive consonants, paired with vowel /a/ in CV context, were added to the list of stimuli, as seed sounds, to expand the entropy pool, to reduce any effect of guessing [4, 13].

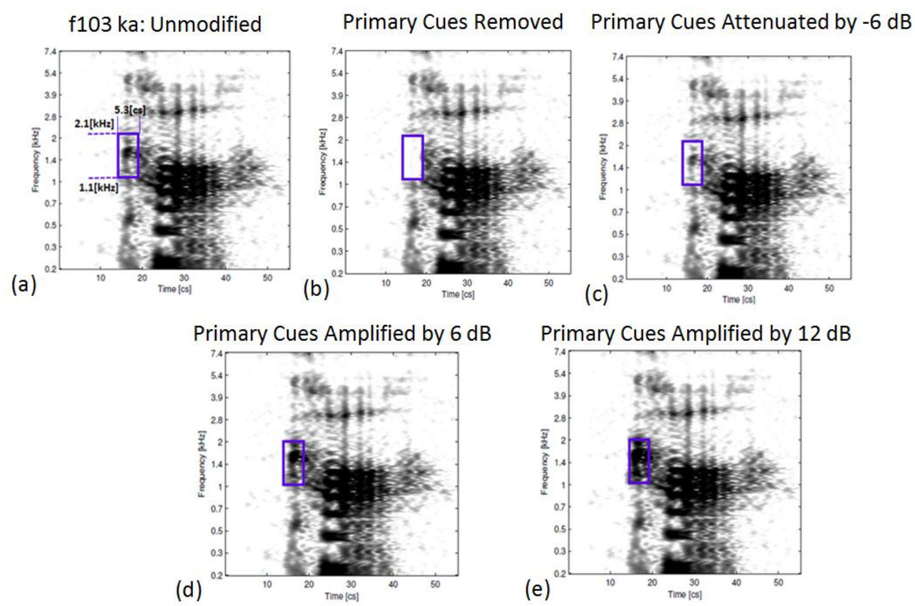


Figure 2.2: *Example of varying the masking of the primary cue for female talker f103 speaking /ka/: (a) Unmodified version. (b) Primary cue removed. (c) Primary cue attenuated by 6 [dB]. (d) Primary cue amplified by 6 [dB]. (e) Primary cue amplified by 12 [dB]. Figure adapted from [4].*

Experiment I: Masking of the primary cue at a token-dependent time-frequency feature region was based on the 3DDS method of [15, 17]. An example of the four types of modifications utilizing the AI-gram [a spectrogram based on human critical bandwidths [17]] of female talker f103 saying /ka/ is shown in Fig. 2.2. The top-left panel (a) is the unmodified version, labeled with the primary cue region [16]. The remainder of the panels have been modified: (b) primary cue removed, (c) primary cue attenuated by 6 dB, and (d, e) primary cue amplified by 6 and 12 dB. Three different wide-band SNRs at 0, 9, 18 [dB] (white noise) were used [4, 13].

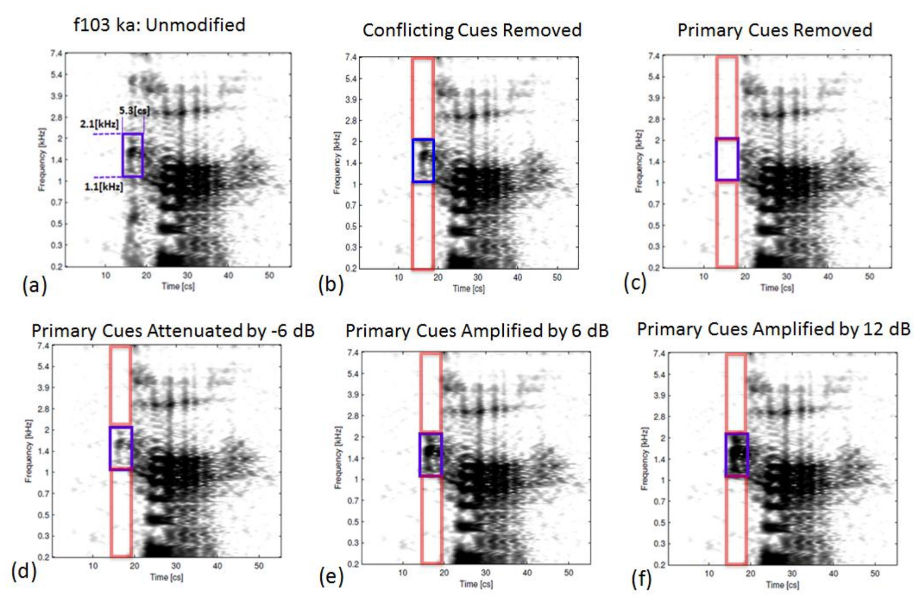


Figure 2.3: *The removal of the conflicting cues for female talker f103 speaking /ka/: (a) Unmodified version. (b) Primary cue unmodified. (c) Primary cue removed. (d) Primary cue attenuated by 6 [dB]. (e) Primary cue amplified by 6 [dB]. (f) Primary cue amplified by 12 [dB]. Figure adapted from [4].*

Experiment II: Experiment II was identical to I except the conflicting cues were removed. Examples of AI-grams with the removal of the conflicting cues for f103 /ka/ are shown in Fig. 2.3. The unmodified version shown in (a) is compared to the modified versions shown in (b)-(f). The areas outlined that are above and below the primary cue (i.e. middle outlined area) are the regions where the conflicting cues are removed. Figure 2.3(b) shows the primary cue unmodified and the conflicting cue removed. The middle outline in Fig. 2.3(c)-(f) is modified similarly to Fig. 2.2(b)-(e). The SNR conditions for Exp II and Exp I are identical [4, 13].

CHAPTER 3

EXPERIMENT I RESULTS

Next we discuss how masking the plosive's primary cues affects a subject's ability to identify consonants and vowels for the ANH and 17 HI ears, measured on the same task. Experiment I (Exp I) looks at the role of the primary cues with the conflicting cues in place. The error for each HI ear is compared to the baseline formed from the ANH error patterns. The results show the nature of the 17 HI errors relative to the ANH ear. The key issue is the idiosyncratic token confusions across different HI ears.

The analysis of the data was performed based on the subject, consonant, talker, modification, and SNR. The experiments are identical for the NH and HI subjects. For the 5 NH subjects, only one ear was measured, as both ears were normal.

The responses were tabulated as 8×4 confusion matrices (CM), with each row labeling by one of the 8 tokens (two talkers × four plosives), and each column recording the reported consonants [19]. Due to the seeds, there were many "other" responses, which were reported in a single fifth column. Since there were 3 SNR, there were a total of 3x8x5 CMs (SNR × tokens × responses) for each of the the 17 HI ears.

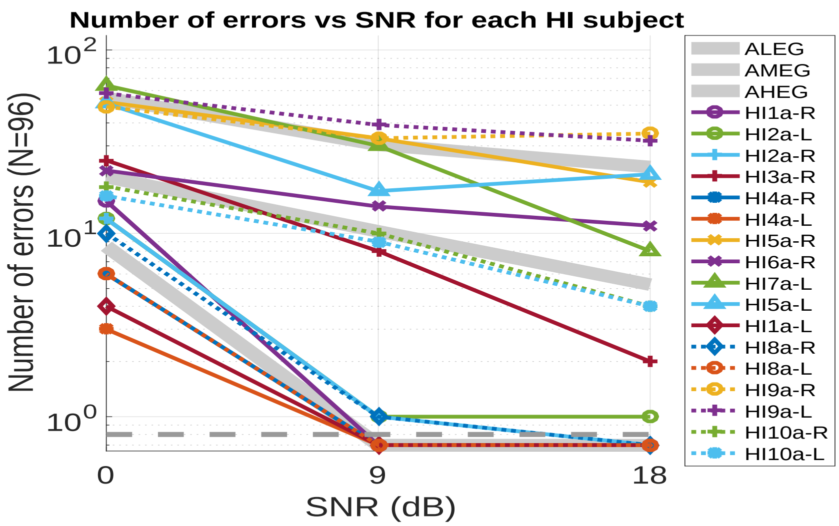


Figure 3.1: *Errors for each HI subject at each SNR. The HI subjects naturally separated into three error groups (i) 60-65, (ii) 16-23, (iii) 3-15 errors, out of a maximum* $N = 96$*, as may be read from the figure at 0 [dB] SNR. The averaged number of errors for each group is represented with the thick gray line. Data points below dashed-gray line indicate zero error.*

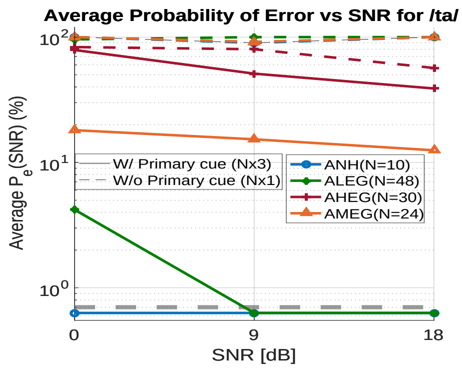
Based on the HI confusion matrices from Exp I, we plot the average error counts for the ears for each SNR condition in Fig. 3.1. The HI subjects split into three groups, as may be seen from the figure: (i) a low error group (LEG), (ii) a medium error group (MEG) and (iii) a high error group (HEG). Subjects from the LEG have zero error at 9 and 18 [dB] SNR on average and between 3-16 errors at 0 [dB] SNR (note: two subjects had 1 error at 9 [dB] SNR, and 1 at 18 [dB] SNR). The LEG overall error shows a strong resemblance to ANH behavior, but with higher error at

0 [dB] SNR. The averages of the three groups are shown as wide solid gray lines.

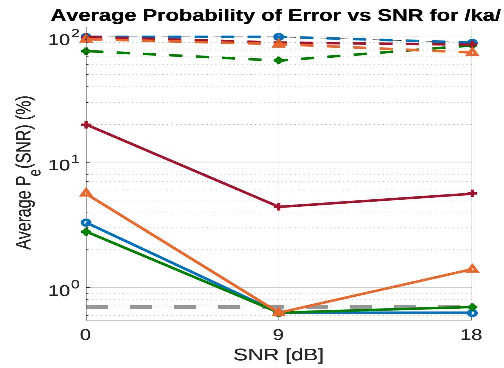
Subjects from the MEG have a much higher error than the LEG at all SNRs. The overall pattern for all subjects is clear, with most errors occurring at 0 [dB] SNR and the fewest occurring at 18 [dB] SNR. Subjects in the HEG, on the other hand, show an idiosyncratic behavior. While all HEG subjects have the most errors at 0 [dB] SNR, not all of them have the fewest error at 18 [dB] SNR. It is not clear if the shallow upturns for ears HI5a-L and HI9a-R are significant, but it seems likely they are. The differences among these three groups shall be further analyzed in the following section.

Next the influence of the primary cue for each of the four consonants is discussed. Figure 3.2 shows the average probability of error as a function of SNR for the four groups. The solid lines indicate the primary cue is presented, while the dashed lines indicate the primary cue has been removed.

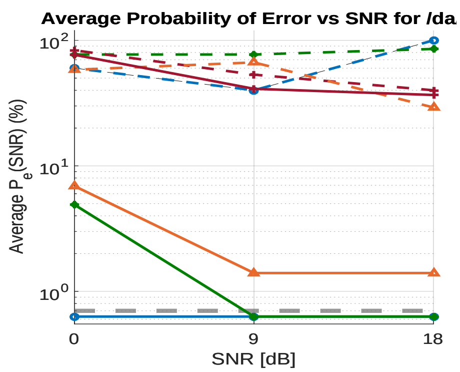
When the primary cue is present, results from the LEG and ANH group are similar. Expectedly the LEG has higher error than the ANH ear, while the MEG and HEG have much higher error for all plosives. Once the primary is removed, the probability of error (dashed lines) for all four groups dramatically increases to more than 80%, showing that the primary cue is crucial for accurate speech recognition. This was known for the NH ears [15], but is the working hypothesis of Exp I for the HI subjects.



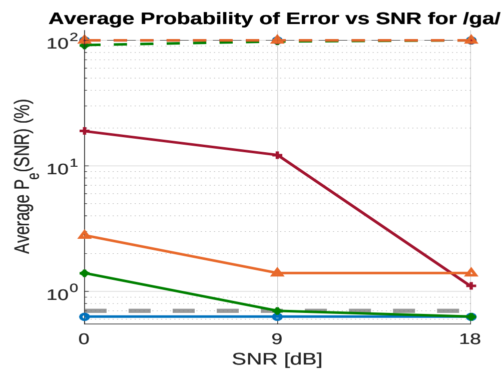
(a) /ta/



(b) /ka/



(c) /da/



(d) /ga/

Figure 3.2: *Average probability of error vs. SNR for /ta/, /ka/, /da/, and /ga/ (legend of (a) applies to all subfigures). Solid line: primary cue presented. Dashed line: primary cue removed. Differences between primary cue present and primary cue removed are large for all subjects. Each symbol corresponds to each group: blue ○: ANH; green ∗: average LEG (ALEG); yellow △: MEG (AMEG); red + average HEG (AHEG). For primary cue present cases, the ANH group has 30 trials, the ALEG has 144 trials, the AMEG has 72 trials, and the AHEG has 90 trials. For primary cue removed cases, each group has one third of the trials in primary cue present cases. Data points below dashed-gray line indicate zero error.*

For /ta/ and /da/, the AHEG (+) shows smaller increase of error compared to those of /ka/ and /ga/.

The explanation may be found from the audiometric profile of the HEG. Both

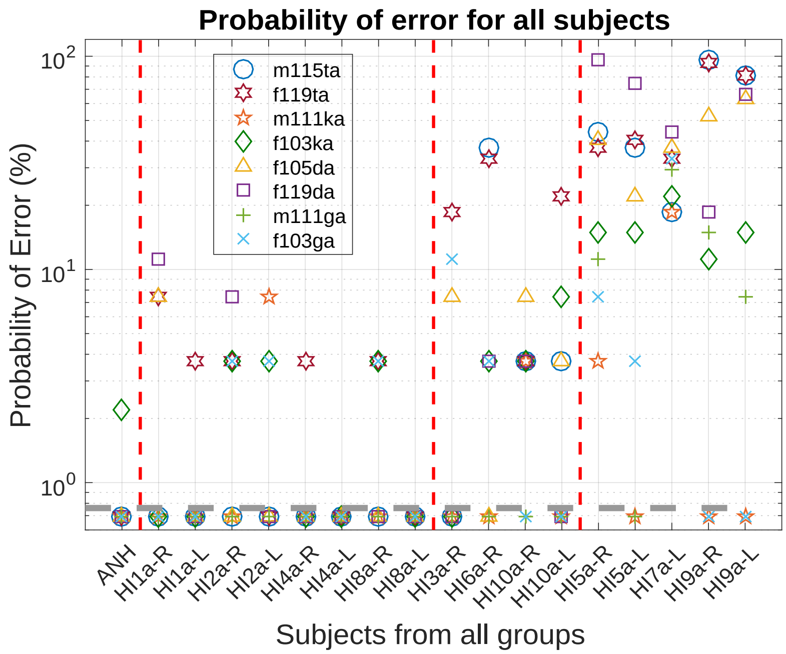


Figure 3.3: *Probability of error of each token for all subjects among four groups: three red vertical dashed lines separate the scatter-plot into four sections indicating four groups: ANH, LEG, MEG, HEG, with increasing error from left to right. For each subject, the probability of error for each token is plotted within its column. There are in total 45 trials for the ANH group, 27 trials for each individual in the LEG, MEG, and HEG. Data points below dashed gray line indicate zero error token.*

/ta/ and /da/ have their primary cues in the high-frequency region, 2–7 kHz, where subjects from the HEG have severe hearing loss. Subjects cannot decode the sound accurately when the primary cues lie in the region where they have the greatest hearing loss. For several HEG ears, the error was close to 100%, thus removal of the primary cue had little impact since removing the primary cue could not result in a significant increase. For all the other cases, the effect of the primary cue is large. Generally speaking, the conclusion is that HI ears are using the same primary cue as NH ears, thus the intensity of the primary cue greatly impacts their speech performance.

Probability of error of all subjects: To study the effect of the subject's hearing loss on consonant recognition, the probability of error for all subjects is plotted in Fig. 3.3. In this way, one can easily compare the behavior among different groups; at the same time one can see how the performance of different tokens changed among the subjects from the same group. As shown in the figure, the scatter-plot is separated into four sections with three vertical dashed lines. The sections correspond to ANH, LEG, MEG, and HEG, from left to right. Individual results for the three HI groups are broken out so that the individual HI behavior can be observed more precisely. One can clearly observe the overall change from group to group. Probability of error for ANH and LEG mainly stays under 10%. MEG shows an increase of error rate where the highest error lies around 40%. The probability of error for HEG is between 10 and 100%. Very few subjects have zero error tokens.

ANH and LEG subjects: The similarity of the LEG and the ANH group subjects was previously mentioned. For the ANH group, there is only one error made by one subject among the five subjects, for token f103ka out of 45 trials. As shown in Fig. 3.3, for the LEG individuals, two subjects, HI4a-L and HI8a-L, had zero error for all tokens and most of the other tokens had zero error. Subjects HI1a-L and HI4a-R made one error for token f119ta, HI8a-R made one error for tokens f119ta, f103ka, and f103ga. HI2a-R and HI2a-L had several single error tokens, and two errors for tokens f119da and m111ka. The worst case was subject HI1a-R who had three non-zero error tokens: two errors were made for f105da and f119ta, while three errors were made for f119da. The probability of error for the LEG subjects is below 11%, with more than half being below 4%.

MEG subjects: The number of zero errors for these subjects dropped dramatically except for HI3a-R, who had the fewest non-zero errors. The remaining MEG subjects had increased errors between 4 and 40 %.

HEG subjects: These subjects had the smallest number of zero-error tokens. HI9a-L made no mistakes for tokens m111ka and f103ga. This makes sense since HI9a has the smallest hearing loss for low frequencies between 0.5 and 2 kHz, where /ka/ and /ga/ have their primary speech cues, in the low to mid frequencies, as shown in Fig. 2.2.

When viewed at the single ear (idiosyncratic) level, there is a logical relation between the hearing loss and the error patterns. Consonants /ta/ and /da/ are both high-frequency plosives, with the burst between 4 and 8 kHz [15]. Subjects from the

HEG all have severe hearing loss in this frequency range, so it is expected that they have high errors for those sounds. HI7a-L, unlike other ears in the HEG, has all tokens in the high error range. This follows from the significant hearing loss level at all frequencies.

In summary, there is a definite relation between the hearing profile and the consonant lost primary cues. It is rare to see such tight relationships. But then most, if not all studies, do not try to correlate the speech cues and the hearing loss in such a detailed (i.e., idiosyncratic) way. In this study, the details of the hearing loss and the speech cues are correlated.

In the next chapter, the effect of removing the conflicting cues will be explored.

CHAPTER 4

EXPERIMENT II RESULTS

Experiment II (Exp II) studies the effect of removing the conflicting cues with the primary cue present. Generally, the same methods were used when processing the data from Exp II, including the same grouping criteria. The numbers of trials for each group are the same as for Exp I.

As previously mentioned, Kapoor and Allen [13] found that for NH ears, the conflicting cues were the dominant source of error in speech perception. By removing conflicting cues, an improvement in consonant scores would be expected. The following experimental results show that this intuition is correct, for some subjects, while other subjects show a more complex behavior.

A comparison of Exp I and Exp II error is shown in Fig. 4.1. The abscissa represents probability of error in Exp I, while the ordinate represents error in Exp II. Each token is indicated by a different symbol. Groups ANH, LEG, MEG and HEG are indicated by different colors. Data points below or on the left side of the dashed gray line have zero error in Exp I or Exp II respectively. If a token falls below the

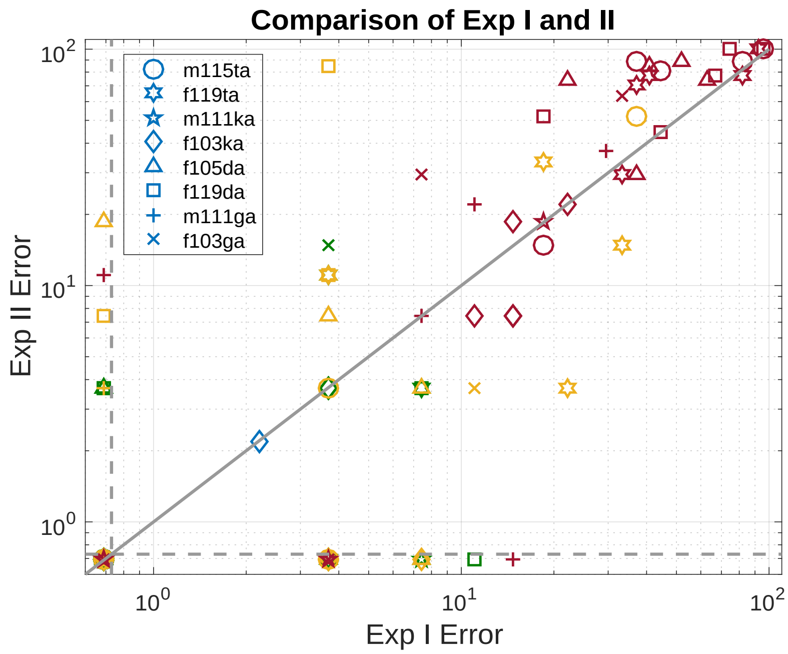


Figure 4.1: *Token scatter-plot of probability of error of Exp I and II: the abscissa indicates Exp I error while the ordinate represents Exp II error. Each token is represented with a different marker shape as shown in the legends. Four colors of the marker indicate four groups: blue: ANH; green: LEG; orange: MEG; red: HEG. Data points above the diagonal line indicate a higher error for Exp II, meaning performance degraded after removing conflicting cues. Data points falling below the diagonal line indicate a lower error for Exp II, meaning performance improved. Data points below the horizontal dashed gray line indicate zero error in Exp I. Data points on the left side of the vertical dashed gray line indicate zero error in Exp II.*

diagonal line, then it has higher error in Exp I than in Exp II, meaning the performance improved after removing the conflicting cues. Points above the diagonal degraded following the removal of conflicting cues.

For the ANH group, seven out of eight tokens had zero error (f103ka had 2.2% error). As shown in Fig. 4.1, all the zero tokens remain zero error for Exp II, as they all fall below and on the left side of the horizontal and vertical zero token reference lines. The one non-zero error token f103ka falls on the diagonal line, indicating no change for NH subjects. These results show that the conflicting cues do not impact the performance of NH subjects, as expected. NH ears only use the primary cue to make their decisions.

For subjects in the LEG and MEG, approximately equal numbers of tokens show improvement and degradation in Exp II. Subjects in the HEG show major degradation, as most HEG tokens reside above the diagonal line. The behavior of each group shall be studied in greater detail next.

LEG behavior: In Fig. 4.1, tokens for the LEG are rather evenly distributed about the diagonal. One interesting token f103ga ($\times$) falls above the diagonal line, with an increase from 3.7% to 14.8% from Exp I to II. Token f119da ($\square$) falls below the diagonal line with zero error for Exp II and 11.1% error for Exp I. There exist two tokens with 3.7% error for Exp II and zero error for Exp I, and two other tokens with 3.7% error for Exp II and 7.4% error for Exp I. Out of the total of 64 points (8 tokens $\times$ 8 individuals), in total 11 data points show improvement (lower error) for Exp II, 9 data points show degradation, and the remaining 40 points remain on the diagonal line. In summary, the removal of conflicting cues has only a minor positive effect on the LEG performance.

MEG behavior: The MEG data points are widely spread. Out of 32 data points (8 tokens $\times$ 4 individuals), 11 points show improvement, 9 data points show degradation, and 12 points remain the same. Thus MEG subjects are more likely to be affected by the removal of conflicting cues than the LEG subjects.

HEG behavior: The large majority of the HEG data points have increased errors when the conflicting cues are removed. From Fig. 4.1, out of 40 data points, 10 points display improvement, 21 data points degrade and 9 data points remain the same. When a token improves, it shows that the subject is using a very different strategy to decode the speech sound. The HEG subjects must be using both primary *and* conflicting cues, since removal of the conflicting cues, on average, reduces the performance.

For example, HEG ear HI7a-L either shows improvement or remains the same for all tokens, with the exception of /ga/ tokens. Thus we conclude that ear HI7a-L is using conflicting cues as an aid to correctly understand sound with consonant /ga/. Unfortunately HI7a-R was not tested. Moreover, as may be seen from Fig. 3.3, HI7a-L is significantly less variable than other HEG subjects (but not the smallest), and is using the conflicting cues for this performance, since the error increases when they are removed.

Another interesting common finding for all HI subjects is that for both /ka/ tokens, the performance is never worse. The most reasonable explanation is that /ka/ is a low-frequency burst and all HI subjects have their best hearing level in the low-frequency region. Their focus may be reinforced by removing high-frequency conflicting cues.

In conclusion, for the ANH group, the conflicting cues are not a source of confusion compared to noise. For the LEG and MEG, the removal of conflicting cues does have a slight positive effect on their overall performance. For the HEG, their behaviors are idiosyncratic. The majority of them appear to be using conflicting cues in addition to the primary cue. For one ear, HI7a-L, the conflicting cues are a major source of confusion for most consonants, except for /ga/. Another theory with some support is that /ga/ has several primary cues, but evidence for this idea is sparse.

CHAPTER 5

ENTROPY ANALYSIS

In this chapter, entropy is introduced to study the idiosyncratic patterns of each token. Entropy is a measure of the size of the confusion group, in other words, a metric of randomness of the error. The entropy vs. error for Exp II is plotted in Fig. 5.1. Such entropy-error relationships were first introduced by [11].

Different groups are indicated by different colors. To distinguish among different subjects for the HEG, different marker sizes are used. From the smallest to the largest marker, each corresponds to subject HI5a-R, HI5a-L, HI7a-L, HI9a-R, and HI9a-L. A separate figure is made to better show the patterns of the HEG entropy, as shown in Fig. 5.2. In the separate figure, each subject is identifiable by a different marker. To further show the results for each subject in the HEG, the error and entropy data are shown in Table 5.1.

For each token, there can exist five outcomes (responses), /ta/, /ka/, /da/, /ga/, and other. Thus the maximum (totally random) entropy is $\log_2(5) = 2.32$ [bits]. When the entropy is 0 bit, the pattern must be completely consistent. The higher

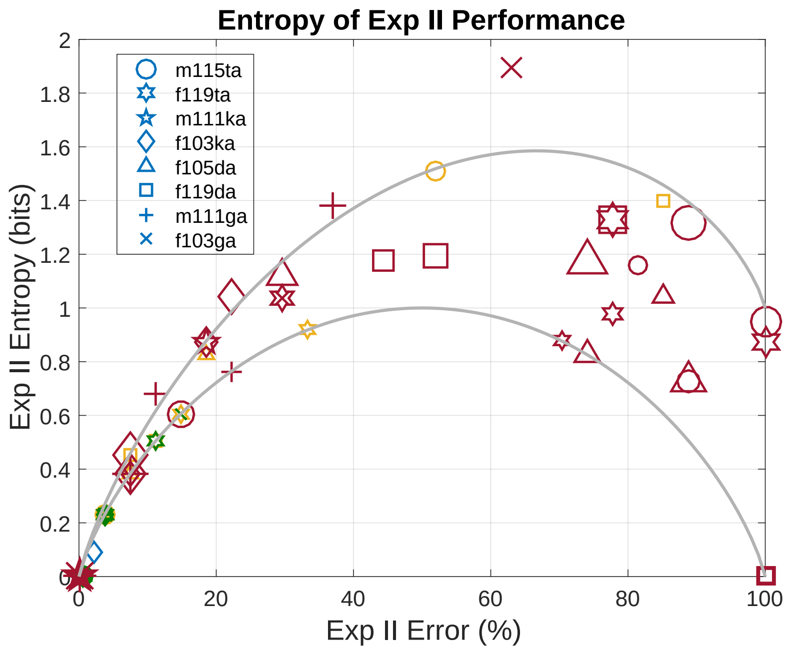


Figure 5.1: *Scatter-plot of entropy analysis of Exp II entropy as a function of error. Each token is represented with a different marker, as shown in the legends. Four colors of the marker indicate four groups. Red, orange, green, blue correspond to the HEG, MEG, LEG, and ANH group, respectively. The majority of the HEG subjects are spread out. Different marker sizes are used for the HEG to emphasize the different subjects. Two gray curves are plotted as references to different entropy bits. Lower curve is when entropy is 1 bit, while higher curve represents 1.585 entropy bits.*

the entropy, the more random the pattern, corresponding to a more idiosyncratic result.

Two reference curves for two (1 bit) and three (1.585 bits) outcomes are summarized in the figures. When a point lies on the 1-bit curve, there are exactly two outcomes, thus their probabilities must add to 1. For points on the 1.585-bit entropy curve, or between two curves, there are typically between two and three outcomes, with two of them having similar probabilities, and the three adding to 1. Tokens above the 1.585-bit entropy curve must have more than three outcomes. Tokens lying below the 1.585-bit line with similar probabilities can, in theory, have four outcomes (but this will be unlikely).

The entropy is a representation of the idiosyncratic behavior among the HEG subjects. Some subjects in the HEG can successfully recognize certain tokens without error. Two 100% error tokens are both f119da from subjects HI5a-R and HI5a-L. All 0% error tokens are low-frequency burst plosives: m111ka, m111ga, and f103ga.

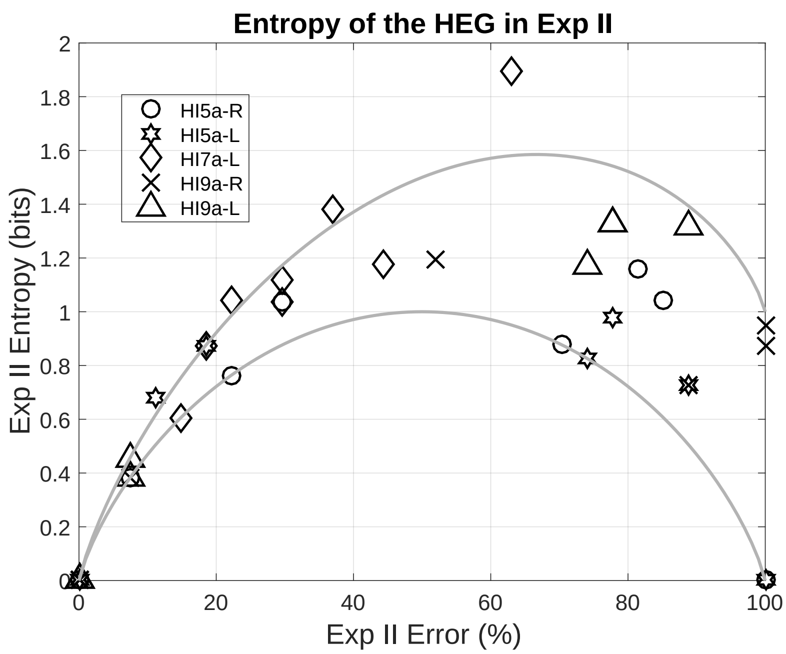


Figure 5.2: *Scatter-plot of entropy analysis of Exp II entropy as a function of error for the HEG. Each subject is represented with a different marker, as shown in the legends. Two gray curves are plotted as references to different entropy bits. Lower curve is when entropy is 1 bit, while higher curve represents 1.585 entropy bits.*

Tokens f103ka, m111ga, f103ga for subject HI7a-L, and token m111ga for HI5a-L, fall above the three-outcome curve. As discussed earlier, the HEG subjects show the most variability. This can also be observed in Fig. 5.1 and Fig. 5.2 since the points from the HEG are the most spread out.

Table 5.1: *Error and entropy data for the HEG (Left: Error (%); Right: Entropy (bits)). These detailed data support Fig. 5.1 and Fig. 5.2. The corresponding ordinates for each point are presented.*

	HI5a-R		HI5a-L		HI7a-L		HI9a-R		HI9a-L	
m115ta	82	1.16	89	0.73	15	0.61	100	0.95	89	1.32
f119ta	70	0.88	78	0.98	30	1.04	100	0.87	78	1.33
m111ka	0	0	0	0	19	0.87	0	0	0	0
f103ka	7	0.38	19	0.87	22	1.04	7	0.38	7	0.46
f105da	85	1.04	74	0.83	30	1.12	89	0.73	74	1.17
f119da	100	0	100	0	44	1.18	52	1.19	78	1.33
m111ga	22	0.76	11	0.68	37	1.38	0	0	7	0.38
f103ga	30	1.04	0	0	63	1.89	0	0	0	0

The above findings can be easily observed from Table 5.1. In each subject column the number on the left is the probability of error [%], and the number on the right is the corresponding entropy [bits]. This table is an alternative representation of Fig. 5.1. As may be seen from the table, the highest entropy is 1.89 bits, for subject HI7a-L, token f103ga, which lies closest to the 2-bit limit in Fig. 5.1. Five outcomes were reported for this case which suggests the subject was guessing.

As observed from Fig. 5.2, a cluster of data points falls in the region with error ranging from 70% to 85% and entropy between 0.98 bit and 1.33 bits. This distribution of errors shows that each subject had either three or four outcomes, and that different combinations of outcomes were reported. Thus, the result shows that different tokens likely sound quite different for different ears, reinforcing the likelihood of idiosyncratic behavior between the HEG ears.

CHAPTER 6

CONCLUSION

In this research, there were two objectives. Exp I examined the effect of the plosive's primary cue in the HI ear. Exp II examined the role of conflicting cues. The analysis from Exp I shows that HI listeners are using the same primary cue as NH ears for correct recognition. The strength of the primary cue is critical for low-error HI speech perception, and is especially important in the presence of noise.

The HI ears were then separated into three error groups: LEG, MEG, and HEG. Overall, the LEG showed results similar to the ANH group, but with slightly higher error, ranging from 3% to 11%. However, subjects in the MEG and HEG show significantly more variability, as their performance depends on the frequency range of their hearing loss.

Subjects in all three HI groups exhibit sensitivity to the presence of conflicting cues, as shown in Figs. 4.1 and 5.1. For the MEG and HEG subjects, the removal of conflicting cues can either enhance or degrade speech perception. Interestingly, a few ears from both groups seem to depend on conflicting cues for correct speech recognition. The experimental removal of the conflicting cues demonstrates that some HEG ears are using not only the primary cue, but also conflicting cues, for correct speech perception.

Errors are dependent on the specific consonants and are highly variable across subjects, demonstrating the idiosyncratic nature of the subjects. Subjects in the MEG and HEG make more mistakes for high-frequency burst consonants, /ta/ and /da/, and fewer mistakes for mid- to low-frequency burst consonants, /ga/ and /ka/.

The results suggest that subjects belonging to the LEG are more likely to benefit from hearing aids. Thus it would be clinically useful to classify subjects in this way, to predict success with their aided fits. Such a classification could also be used to improve the fitting algorithm.

REFERENCES

[1] J. B. Allen, "Derecruitment by multiband compression in hearing aids," in *Psychoacoustics, Speech, and Hearing Aids*, B. Kollmeier, Ed. Singapore: World Scientific Press, 1996, pp. 141–152.

[2] J. B. Allen, "Amplitude compression in hearing aids," in *MIT Encyclopedia of Communication Disorders*, R. Kent, Ed. MIT, Boston Ma: MIT Press, 2003, ch. Part IV, pp. 413–423.

[3] H. Dillon, "Measuring hearing aids in couplers and ear simulators," *Hearing Aids*, 2012.

[4] C. L. Cole, "On the effects of masking of perceptual cues in hearing-impaired ears," Ph.D. dissertation, University of Illinois, Urbana-Champaign, 2017.

[5] A. Abavisani and J. B. Allen, "Evaluating hearing aid amplification using idiosyncratic consonant errors," *J. Acoust. Soc. Am.*, vol. 142, no. 6, pp. 3736–3745, 2017.

[6] P. M. Zurek and L. A. Delhorne, "Consonant reception in noise by listeners with mild and moderate sensorineural h earing impairment," *JASA*, vol. 82, no. 5, pp. 1548–1559, 1987.

[7] A. Trevino and J. B. Allen, "Individual variability of hearing impaired consonant perception," *Seminars in Hearing*, vol. 34, no. 2, pp. 74–85, 2013, guest editor: Jason Gaalster, PhD.

[8] A. Trevino and J. B. Allen, "Within-consonant perceptual differences in the hearing impaired ear," *J. Acoust. Soc. Am.*, vol. 134(1), pp. 607–617, 2013.

[9] S. Phatak, A. Lovitt, and J. B. Allen, "Consonant confusions in white noise," *J. Acoust. Soc. Am.*, vol. 124, no. 2, pp. 1220–33, 2008.

[10] S. A. Phatak, Y. Yoon, D. M. Gooler, and J. B. Allen, "Consonant loss profiles in hearing impaired listeners," *J. Acoust. Soc. Am.*, vol. 126, no. 5, pp. 2683–2694, 2009.

[11] R. Singh and J. B. Allen, "The influence of stop consonants' perceptual features on the Articulation Index model," *J. Acoust. Soc. Am.*, vol. 131, no. 4, pp. 3051–3068, 2012.

[12] J. Toscano and J. B. Allen, "Across and within consonant errors for isolated syllables in noise," *Jol. of Speech, Language, and Hearing Research*, vol. 57, no. DOI: 10.1044/2014_JSLHR-H-13-0244, pp. 2293–2307, 2014.

[13] A. Kapoor and J. B. Allen, "Perceptual effects of plosive feature modification," *J. Acoust. Soc. Am.*, vol. 131, no. 1, pp. 478–491, 2012.

[14] Y. Yoon, J. Allen, and D. Gooler, "Relationship between consonant recognition in noise and hearing threshold," *J. Speech, Language, and Hearing Research,*, vol. doi: 10.1044/1092-4388(2011/10-0239), pp. 460–473, 2012.

[15] F. Li and J. B. Allen, "Manipulation of consonants in natural speech," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 19, no. 3, pp. 496–504, 2011.

[16] A. Menon, F. Li, and J. B. Allen, "A psychoacoustic method to find the perceptual cues of stop consonants in natural speech," *J. Acoust. Soc. Am.*, vol. 127(4), pp. 2599–2610, 2010.

[17] J. B. Allen and F. Li, "Speech perception and cochlear signal processing," *IEEE Signal Processing Magazine*, vol. 26, no. 4, pp. 73–77, 2009, invited.

[18] S. Phatak and J. B. Allen, "Consonant and vowel confusions in speech-weighted noise," *J. Acoust. Soc. Am.*, vol. 121, no. 4, pp. 2312–26, 2007.

[19] J. B. Allen, *Articulation and Intelligibility.* Morgan and Claypool, 2005.